

**Notice of References Cited**Application/Control No.  
10/055,586Applicant(s)/Patent Under  
Reexamination  
BAR-YOSSEF ET AL.Examiner  
Luke S. WassumArt Unit  
2177

Page 1 of 5

**U.S. PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-5,909,677 A	06-1999	Broder et al.	707/3
	B	US-6,119,124 A	09-2000	Broder et al.	707/103R
	C	US-6,138,113 A	10-2000	Dean et al.	707/2
	D	US-6,230,155 B1	05-2001	Broder et al.	707/3
	E	US-6,349,296 B1	02-2002	Broder et al.	707/3
	F	US-6,614,764 B1	09-2003	Rodeheffer et al.	370/254
	G	US-6,615,209 B1	09-2003	Gomes et al.	707/5
	H	US-6,658,423 B1	12-2003	Pugh et al.	707/102
	I	US-6,665,837 B1	12-2003	Dean et al.	715/501.1
	J	US-			
	K	US-			
	L	US-			
	M	US-			

**FOREIGN PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

**NON-PATENT DOCUMENTS**

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Manber, U. "Finding Similar Files in a Large File System", Technical Report TR 93-33, University of Arizona, Department of Computer Science, October 1993.
	V	Broder, A.Z. "Some Applications of Rabin's Fingerprinting Method", in R. Capocelli, A. De Santis, U. Vaccaro (eds), "Sequence II: Methods in Communications, Security and Computer Science", Springer-Verlag, 1993.
	W	Agrawal, R. and R. Srikant "Fast Algorithms for Mining Association Rules", Proceedings of the 20 <sup>th</sup> VLDB Conference, pp. 487-499, 1994.
	X	Brin, S., J. Davis and H. Garcia-Molina "Copy Detection Mechanisms for Digital Documents", Proceedings of the ACM SIGMOD Conference, pp. 398-409, May 1995.

\*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)  
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

<b>Notice of References Cited</b>	Application/Control No. 10/055,586	Applicant(s)/Patent Under Reexamination BAR-YOSSEF ET AL.	
	Examiner Luke S. Wassum	Art Unit 2177	Page 2 of 5

**U.S. PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

**FOREIGN PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

**NON-PATENT DOCUMENTS**

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Heintze, N. "Scalable Document Fingerprinting (Extended Abstract)", Proceedings of the 1996 USENIX Workshop on Electronic Commerce, November 1996.
	V	Broder, A.Z. "On the Resemblance and Containment of Documents", Proceedings of Compression and Complexity of SEQUENCES, p. 21, 11-13 June 1997.
	W	Broder, A.Z., S.C. Glassman, M.S. Manasse and G. Zweig "Syntactic Clustering of the Web", Proceedings of the 6 <sup>th</sup> International World Wide Web (WWW) Conference (WWW6), pp. 1157-1166, 1997.
	X	Fang, M., N. Shivakumar, H. Garcia-Molina, R. Motwani and J.D. Ullman "Computing Iceberg Queries Effectively", Proceeding of the 24 <sup>th</sup> VLDB Conference, 1998.

\*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)  
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

**Notice of References Cited**

Application/Control No.

10/055,586

Applicant(s)/Patent Under  
Reexamination  
BAR-YOSSEF ET AL.

Examiner

Luke S. Wassum

Art Unit

2177

Page 3 of 5

**U.S. PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

**FOREIGN PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

**NON-PATENT DOCUMENTS**

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Kumar, R., P. Raghavan, R. Rajagopalan and A. Tomkins "Trawling the Web for Emerging Cyber-Communities", Proceedings the 8 <sup>th</sup> International World Wide Web (WWW) Conference (WWW8), pp. 1481-1493, 1999.
	V	W3C "Document Object Model (DOM) Level 2 Core Specification Version 1.0, W3C Recommendation 13 November 2000", downloaded from www.w3.org.
	W	Davidson, B.D. "Recognizing Nepotistic Links on the Web", Proceedings of the AAAI-2000 Workshop on Artificial Intelligence for Web Search, pp. 23-28, 2000.
	X	Chakrabarti, S., M. Joshi and V. Tawde "Enhanced Topic Distillation Using Text, Markup Tags and Hyperlinks", Proceedings of the ACM SIGIR Conference, 9-12 September 2001.

\*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)

Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

**Notice of References Cited**

Application/Control No.

10/055,586

Applicant(s)/Patent Under  
Reexamination  
BAR-YOSSEF ET AL.

Examiner

Luke S. Wassum

Art Unit

2177

Page 4 of 5

**U.S. PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

**FOREIGN PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

**NON-PATENT DOCUMENTS**

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Crescenzi, V., G. Mecca and P. Merialdo "RoadRunner: Towards Automatic Data Extraction from Large Web Sites", Proceedings of the 27 <sup>th</sup> VLDB Conference, 2001.
	V	Bar-Yossef, Z. and S. Rajagopalan "Template Detection via Data Mining and its Applications", Proceedings of the WWW2002 Conference, pp. 580-591, 7-11 May 2002.
	W	Haveliwala, T.H., A. Gionis, D. Klein and P. Indyk "Evaluating Strategies for Similarity Search on the Web", Proceedings of the WWW2002 Conference, 7-11 May 2002.
	X	Crescenzi, V., G. Mecca and P. Merialdo "RoadRunner: Automatic Data Extraction from Data-Intensive Web Sites", Proceedings of the ACM SIGMOD Conference, p. 624, 4-6 June 2002.

\*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)  
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

<b>Notice of References Cited</b>	Application/Control No. 10/055,586		Applicant(s)/Patent Under Reexamination BAR-YOSSEF ET AL.	
	Examiner Luke S. Wassum		Art Unit 2177	Page 5 of 5

**U.S. PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

**FOREIGN PATENT DOCUMENTS**

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

**NON-PATENT DOCUMENTS**

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Laender, A.H.F., B.A. Ribeiro-Neto, A.S. da Silva and J.S. Teixeira "A Brief Survey of Web Data Extraction Tools", SIGMOD Record, Vol. 31, No. 2, pp. 84-93, June 2002.
	V	Arasu, A. and H. Garcia-Molina "Extracting Structured Data from Web Pages", Proceedings of the ACM SIGMOD Conference, 9-12 June 2003.
	W	Yi, L., B. Liu and X. Li "Eliminating Noisy Information in Web Pages for Data Mining", Proceedings of the ACM SIGKDD Conference, 24-27 August 2003.
	X	Ma, L., N. Goharian, A. Chowdhury and M. Chung "Extracting Unstructured Data from Template Generated Web Documents", Proceedings of the 12 <sup>th</sup> International Conference on Information and Knowledge Management, pp. 512-515, 3-8 November 2003.

\*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)  
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.